UNITED STATES PATENT AND TRADEMARK OFFICE

COMMISSIONER FOR PATENTS
UNITED STATES PATENT AND TRADEMARK OFFICE
P.O. BOX 1450
ALEXANDRIA, VA 22313-1450
www.uspto.gov

Fax Cover Sheet

Date: Feb 23, '04

To: Barbara C. miel	**From:** TADESSE HAILU
Application/Control Number: 09/369,360	**Art Unit:** 2173
Fax No.: (312) 258-5921	**Phone No.:** (703) 306-2799
Voice No.: (312) 258-4538	**Return Fax No.:** (703) 872-9306
Re: The missing page #62	**CC:**

☐ **Urgent** ☐ **For Review** ☐ **For Comment** ☐ **For Reply** ☒ **Per Your Request**

Comments:

The missing page #62 from "Summarizing Similarities and Differences Among Related Documents" is enclosed.

Number of pages __ including this page 2

Table 4. Summaries versus Full-Text: Task Accuracy, Time, and User Feedback

Metric	Full-Text	Summary
Accuracy (Precision, Recall)	30.25, 41.25	25.75, 48.75
Time (mins)	**24.65**	**21.65**
Usefulness of text in deciding relevance (0 to 1)	.7	.8
Usefulness of text in deciding irrelevance (0 to 1)	.7	.6
Preference for more or less text	"Too Much Text."	"Just Right."

collection of pairs of articles on international events culled from searches on the World Wide Web, including articles from Reuters, Associated Press, the Washington Post, and the New York Times. Pairs were selected such that each member of a pair was closely related to the other, but by no means identical; the pairs were drawn from different geopolitical regions so that no pair was similar to another. In the Peru pair only the precision of the top ten sentence pairs is calculated. For the other pairs precision is calculated for all output sentence pairs (on average 50 sentence pairs for Evangelist and 60 for Chechnya). For each document pair the assigned weighting method was applied to each text and the single best match for each sentence was output. The goal of this experiment was to measure the ability of the alignment method to find correct alignments (those that are both correctly aligned and relevant to the user's given topic). Alignment correctness was determined by a human judge.

In Table 3, we see that all of the reweighting schemes outperform the baseline tf.idf measure for these tasks and that the highest average results are obtained with the method which uses spreading and clipping. The results with spreading alone (SPREAD) were also better on average than tf.idf (RAW) with the greatest difference on the Evangelist pair, but small differences on the other pairs. The removal of words using clipping resulting in improvements (on average) for the RAW and SPREAD based methods, but not for the RAWPOL. Clipping results in the most reduction when the differences between minimum and maximum word weights is greatest. This suggests that the proper name weight increment in RAWPOL may have been too large, causing more words, and sometimes useful words, to be removed. These results are only suggestive; conclusive results would require experimenting with a much larger data sample.

9.3. *Effectiveness of Spreading Activation*

In addition to the intrinsic evaluation of alignments, we also carried out an extrinsic evaluation, where we evaluated the usefulness of spreading in the context of an information retrieval task. In this experiment, subjects were informed only that they were involved in a timed information retrieval research experiment. In each run, a subject was presented with a pair of query and document, and asked to determine whether the document was relevant or irrelevant to the query. In one experimental condition the document shown was the full text, in the other the document shown was a summary generated with the top 5 weighted sentences. Subjects (four altogether) were rotated across experimental conditions, but no subject was in both conditions for the same query-document pair. We hypothesized that if the summarization was useful, it would result in savings in time, without significant loss in accuracy.